Case Study 3: Investigating Biomarkers of Kidney Function II

Kyly Hiatt, Miriam Hu, Kenneth Lee, Robert Tumasian, and Sofia Vega

July 19, 2018

1 Introduction

The function of the kidneys is to filter waste from the bloodstream and dispel it from the body as urine. Chronic kidney disease (CKD) is an acute decrease in kidney functioning over time. Without proper filtration, natural waste begins to accumulate in the blood. This inability to properly filter the blood can result in nerve damage, high blood pressure, anemia, and many other damaging symptoms. Frequent causes of kidney disease include physical injury, diabetes, infection, and high blood pressure.

Physicians measure kidney function via glomerular filtration rate, they calculate GFR to categorize CKD severity into six stages (1, 2, 3A, 3B, 4, 5), with one being the least severe. Researchers aim to identify a biomarker that can accurately predict kidney decline to identify high risk patients early on. Previous investigation suggests that soluble urokinase-type plasminogen activator receptor (suPAR) may be a reliable predictor of CKD stage progression, but exploration of baseline GFR, C-reactive protein, and diabetes can contribute to better predictions of patient kidney function. Improved predictions will ensure that the most high risk patients receive treatment to either slow the progression of kidney decline or identify them as priorities to receive a more extreme treatment.

Extreme treatment options, including dialysis and/or transplantation, are not only stressful and exhausting for patients, but they are extremely costly to both patients and medical facilities. Dialysis treatment can cost 70,000 USD a year and kidney transplantation can cost 100,000 USD[1]. Minimizing patients who need drastic treatment will save money for patients and hospitals and improve quality of life.

To understand longterm survival outcomes, this investigation will also analyze the relationship between clinically diagnosed kidney disease patients and major cardiovascular events, as high blood pressure is a common symptom of kidney function decline. The role of additional covariates in terms of survival for these patients will also be explored. A more holistic understanding of contributing factors to kidney decline will help physicians more effectively treat patients and improve survival rates. Table 1 provides several characteristics of the patients in our study.

2 Methods

2.1 Objective #1

We will determine if the associations explored in Case Study #2 simultaneously depend on diabetes, baseline GFR, and C-reactive protein. In order to begin our investigation, we first created new variables to represent the

[1] Tuller, D. Lifesaving Kidney Treatment, but only to a point. *New York Times.* 2009 Mar 12 [accessed 2018 Jul 13]

difference in GFR (follow-up GFR − baseline GFR), and CKD stage at baseline and follow-up. Next, we built two multiple linear regression models, including our predictors of interest and all interaction terms, with baseline GFR and GFR difference as our outcomes. We also built a logistic regression model, including our predictors of interest and all interaction terms, with CKD stage progression as our outcome. We analyzed p-values at the 95% CL to determine which predictors and interactions significantly contributed to predicting the odds of our outcomes. Then, we built reduced models including only the predictors of interest and significant interaction terms. Finally, we compared our linear models by assessing their respective R^2 values.

2.2 Objective #2

Our next goal is to determine if suPAR level is associated with the risk of developing more advanced CKD after adjusting for covariates among those with normal or mild CKD. First, we created a new binary variable to indicate if CKD stage progression occurred at follow-up. Then, we built a logistic regression model by *a priori* determining which covariates were clinically important and included them in our model. According to the National Kidney Foundation, "[t]he two main causes of chronic kidney disease are diabetes and high blood pressure"[2]. Additionally, the NKF states that older people and those racially predisposed to have diabetes and high blood pressure are at higher risk for CKD. Thus, we chose suPAR, diabetes, hypertension, age, and race as our predictors in our logistic regression model. We analyzed p-values at the 95% CL to determine which predictors significantly contributed to predicting the odds of CKD stage progression. We also assessed odds ratios and their respective confidence intervals to determine the effects of our predictors on CKD stage progression.

2.3 Objective #3

Next, we identified potential differences in survival between two groups for six conditions: CKD at baseline, sex, race, history of smoking, history of myocardial infarction, and hypertension. To do this, we created Kaplan-Meier survival plots for time to death, stratified by each of the six predictors. Then, we conducted log-rank tests and assessed p-values at the 95% CL to determine if there are any significant differences in survival between the two groups for every condition.

2.4 Objective #4

Our last goal was to determine the risk of experiencing a major adverse cardiovascular event (MACE) among patients with and without clinical CKD at baseline after adjusting for other covariates. We created a Cox proportional hazards model for time to MACE with baseline CKD, suPAR, hypertension, dyslipidemia, diabetes, BMI, history of smoking, and history of myocardial infarction as our predictors. These covariates are listed as risks in the HEART score, a score used in the emergency department to decide whether to treat patients with chest pain early[3]. suPAR was also included as a biomarker for CKD to determine the relationship between suPAR and

[2] About Chronic Kidney Disease. National Kidney Foundation. n.d. [accessed 2018 Jul 13]

[3] Six, A. J., Backus, B. E., & Kelder, J. C. (2008). Chest pain in the emergency room: value of the HEART score. *Netherlands Heart Journal, 16*(6), 191–196.

time to MACE. We verified the proportional hazards assumption for the model by confirming the Schoenfeld residuals were uncorrelated with the ranked time to MACE for every covariate. Then, we analyzed p-values at the 95% CL to determine which predictors significantly contributed to predicting the hazard ratios for MACE. We also assessed the confidence intervals of the hazard ratios.

3 Results

3.1 Objective #1

In Case Study #2, we observed that suPAR levels significantly predict baseline GFR. Here, we added diabetes, baseline CKD, C-reactive protein and all corresponding interaction terms to a multiple linear regression model. Interaction terms that were not significant were removed from the initial model, improving our adjusted R^2 values from 0.684 (full model) to 0.688 (reduced model). Furthermore, we observed that suPAR ($p < 0.001$), Baseline CKD ($p < 0.001$), and the interaction between suPAR and Baseline CKD ($p < 0.001$) were significant. This interaction suggests that there is a significant change in the relationship between baseline GFR and suPAR depending on the presence or absence of CKD at baseline. A summary of our full and reduced linear models are provided in Tables 2 and 3, respectively.

We also previously observed that suPAR levels did not predict ΔGFR. We added diabetes, baseline CKD, C-reactive protein and all corresponding interaction terms to the model. Interaction terms that were not significant were removed from the model, improving our adjusted R^2 values from 0.127 (full model) to 0.130 (reduced model). Furthermore, we observed that suPAR ($p = 0.048$), diabetes ($p = 0.022$), and the interaction between suPAR and baseline CKD ($p = 0.007$) were significant. A summary of our full and reduced linear models are provided in Tables 4 and 5, respectively.

Additionally, we previously observed a significant relationship between CKD stage progression and suPAR using logistic regression. We added diabetes, BL_CKD, CRP, and all interaction terms to the model. Interaction terms that were not significant were then removed. In our reduced model, suPAR ($p = 0.015$) and baseline CKD ($p = 0.025$) were significant. The odds ratio for suPAR was 1.0002. If we consider 1,000 pg/mL to be the smallest clinically meaningful increase in suPAR, then for an increase of 1,000 pg/mL of suPAR, the odds of CKD stage progression occurring increases by a factor of $(1.0002)^{1000} = 1.22$. A summary of our full and reduced logistic models are provided in Tables 6 and 7, respectively.

3.2 Objective #2

After performing logistic regression, suPAR was the only statistically significant variable ($p = 0.0081$). Thus, we reject the null hypothesis and conclude that there is a relationship between suPAR and the odds of CKD stage progression. The odds ratio for a 1 pg/mL increase in suPAR, holding all covariates constant, is 1.0003. If we consider 1,000 pg/mL to be the smallest clinically meaningful increase in suPAR, then for an increase of 1,000 pg/mL of suPAR, the odds of stage progression occurring increases by a factor of $(1.0003)^{1000} = 1.35$. A summary of our logistic model is provided in Table 8.

3.3 Objective #3

After conducting log-rank tests, we found that only clinical CKD at baseline resulted in significant differences in survival between the two groups. Patients with clinical CKD have lower survival rates than patients without clinical CKD at baseline. For each of the other conditions, the p-values for the log rank test are greater than 0.05, as shown in Table 9. Thus, we fail to reject the null hypothesis that there are differences in survival between the two groups at any time point for the other five conditions.

3.4 Objective #4

For all of the predictors in the Cox proportional hazards model, the hazard ratios were statistically insignificant, as shown in Table 10. We fail to reject all null hypotheses that the hazard ratio of MACE is constant for each unit increase in the predictors. We are 95% confident that the hazard rate of experiencing MACE is between 0.95 and 1.42 times higher for those with clinical CKD at baseline than for those without, holding all other predictors constant. According to Table 11, none of the Schoenfeld residuals of the covariates are correlated with the ranked time to MACE. Thus, the proportional hazards assumption is satisfied.

4 Conclusions

When predicting baseline GFR using suPAR, diabetes, baseline CKD, CRP, and significant 2-way interaction terms, we found suPAR, baseline, and the interaction between suPAR and baseline CKD to be significant. In the model predicting the change in GFR using suPAR, diabetes, baseline CKD, CRP, and significant 2-way interaction terms, we found suPAR, DM, and the interaction between suPAR and baseline CKD to be significant. For the logistic regression with CKD stage progression as the outcome and suPAR as the main predictor of interest (adjusting for diabetes, baseline CKD, and CRP) we found that CKD stage progression and suPAR are associated. For a 1,000 pg/mL increase in suPAR, the odds of progression increases by 22% ($p = 0.0149$). Among patients without clinical CKD at baseline, suPAR is associated with the risk of CKD stage progression after adjusting for covariates. For a 1,000 pg/mL increase in suPAR, the odds of progression at follow-up for those without clinical CKD at baseline increases by 35% ($p = 0.0081$).

When identifying potential differences in survival between the two groups for the six conditions in Objective 3, we can observe that the Kaplan-Meier survival plots for time to death appear very similar between groups for every condition. After using log rank tests to determine if there differences in survival between the two groups, we found there are no statistically significant differences between the two groups at any time point for any of the six conditions. Furthermore, after adjusting for other covariates, baseline clinical CKD was statistically insignificant as a predictor for hazard ratio of MACE. We are 95% confident that the hazard rate of experiencing MACE is between 0.95 times lower to 1.42 times higher for those with clinical CKD at baseline than for those without.

5 Appendix

Table 1. Summary of Patient Characteristics.

Characteristic	Participants (n=366)
GFR (mL/min/1.73m^2)	
Baseline	67.37 ± 23.41
Follow-up	64.79 ± 22.53
suPAR (pg/mL)	3635 ± 1643
C-reactive protein (mg/dL)	5.48 ± 8.98
Creatinine (mg/dL)	1.36 ± 1.54
Age (yr)	63.17 ± 12.40
Body mass index (kg/m^2)	29.56 ± 6.59
Sex (%)	
Male	232 (63.4)
Female	134 (36.6)
Race (%)	
Black	74 (20.2)
Non-Black	292 (79.8)
Myocardial infarction (%)	
History	93 (25.4)
No history	265 (72.4)
Not applicable	8 (2.2)
Diabetes (%)	
Diagnosed	133 (36.3)
Not diagnosed	227 (62)
Not applicable	6 (1.7)
Hypertension (%)	
Diagnosed	265 (71.6)
Not diagnosed	97 (26.5)
Not applicable	7 (1.9)
Dyslipidemia (%)	
Diagnosed	243 (66.4)
Not diagnosed	115 (31.4)
Not applicable	8 (2.2)
Proteinuria (%)	
Diagnosed	22 (6)
Not diagnosed	344 (94)
Smoking (%)	
History	199 (54.4)
No history	157 (42.9)
Not applicable	10 (2.7)
Coronary heart disease (%)	
Diagnosed	226 (61.7)
Not diagnosed	124 (33.9)
Not applicable	16 (4.4)
Use of ACE inhibitors and/or ARBs (%)	
History	253 (69.1)
No history	110 (30.1)
Not applicable	3 (0.8)
Death (%)	
Censored	100 (27.3)
Uncensored	260 (71)
Not applicable	6 (1.6)
Time to Death (days)	2330 ± 1156
Myocardial Infarction (%)	
Censored	335 (91.5)
Uncensored	25 (6.8)
Not applicable	6 (1.6)
Time to Myocardial Infarction (days)	2281 ± 1189
MACE (%)	
Censored	177 (48.4)
Uncensored	183 (50)
Not applicable	6 (1.6)
Time to MACE (days)	1817 ± 1330

Table 2. Full Linear Model of the Relationship between suPAR and Baseline GFR with Diabetes and C-reactive Protein.

Predictor	Estimate	SE	t-value	p-value
(Intercept)	84.384	2.771	30.45	< 0.001*
suPAR	-0.001	0.001	-1.54	0.124
DM	5.871	3.651	1.61	0.109
BL_CKD	-20.778	4.08	-5.09	< 0.001*
CRP	0.045	0.192	0.23	0.815
suPAR:DM	-0.001	0.001	-0.75	0.452
suPAR:BL_CKD	-0.003	0.001	-3.37	0.001*
suPAR:CRP	-0.000	0.000	-0.37	0.709
DM:BL_CKD	0.426	3.452	0.12	0.902
DM:CRP	0.003	0.192	0.02	0.988
BL_CKD:CRP	-0.008	0.212	-0.04	0.969

Table 3. Reduced Linear Model of the Relationship between suPAR and Baseline GFR with Diabetes and C-reactive Protein.

Predictor	Estimate	SE	t-value	p-value
(Intercept)	85.828	2.298	37.35	< 0.001*
suPAR	-0.002	0.001	-2.59	0.010*
DM	3.268	1.484	2.20	0.028*
BL_CKD	-19.851	3.842	-5.17	< 0.001*
CRP	-0.056	0.079	-0.71	0.478
suPAR:BL_CKD	-0.004	0.001	-3.84	< 0.001*

Table 4. Full Linear Model to Predict Decline in Kidney Function with Baseline suPAR, Diabetes, Baseline GFR, and C-reactive Protein.

Predictor	Estimate	SE	t-value	p-value
(Intercept)	-1.193	3.455	-0.35	0.730
suPAR	-0.001	0.001	-0.92	0.356
DM	0.606	4.553	0.13	0.894
BL_CKD	-0.667	5.087	-0.13	0.896
CRP	-0.205	0.240	-0.85	0.394
suPAR:DM	-0.001	0.001	-0.99	0.321
suPAR:BL_CKD	0.003	0.001	2.67	0.008*
suPAR:CRP	-0.000	< 0.001	-0.18	0.861
DM:BL_CKD	-3.016	4.305	-0.70	0.484
DM:CRP	0.115	0.2398	0.48	0.632
BL_CKD:CRP	0.221	0.264	0.84	0.404

Table 5. Reduced Linear Model to Predict Decline in Kidney Function with Baseline suPAR, Diabetes, Baseline GFR, and C-reactive Protein.

Predictor	Estimate	SE	t-value	p-value
(Intercept)	0.525	2.875	0.18	0.855
suPAR	-0.002	0.001	-1.99	0.048*
DM	-4.273	1.857	-2.30	0.022*
BL_CKD	0.397	4.808	0.08	0.934
CRP	-0.073	0.098	-0.75	0.457
suPAR:BL_CKD	0.003	0.001	2.70	0.007*

Table 6. Full Logistic Model to Predict Risk of CKD Progression with Baseline suPAR, Diabetes, Baseline GFR, and C-reactive Protein.

Predictor	Odds Ratio	95% CI	p-value
(Intercept)	0.096	(0.033, 0.254)	< 0.001*
suPAR	1.0004	(1.0001, 1.0007)	0.011*
DM	1.813	(0.508, 6.384)	0.355
BL_CKD	0.782	(0.169, 3.406)	0.748
CRP	1.054	(0.986, 1.152)	0.157
suPAR:DM	0.9999	(0.9995, 1)	0.458
suPAR:BL_CKD	0.9998	(0.9995, 1)	0.334
suPAR:CRP	1	(1, 1)	0.383
DM:BL_CKD	1.487	(0.452, 4.773)	0.507
DM:CRP	0.986	(0.9255, 1.047)	0.639
BL_CKD:CRP	1.016	(0.948, 1.087)	0.652

Table 7. Reduced Logistic Model to Predict Risk of CKD Progression with Baseline suPAR, Diabetes, Baseline GFR, and C-reactive Protein.

Predictor	Odds Ratio	95% CI	p-value
Intercept	0.191	(0.106, 0.338)	< 0.001*
suPAR	1.0002	(1.00004, 1.0004)	0.015*
DM	1.17	(0.705, 1.928)	0.540
BL_CKD	0.52	(0.289, 0.912)	0.025*
CRP	1.01	(0.985, 1.037)	0.429

Table 8. Logistic Model to Determine the Association between Covariates and CKD Progression among those with Normal or Mild CKD.

Predictor	Odds Ratio	95% CI	p-value
(Intercept)	0.091	(0.016, 0.461)	0.005*
suPAR	1.0003	(1.000, 1.001)	0.008*
DM	0.994	(0.529, 1.838)	0.984
HTN	1.937	(0.957, 4.122)	0.074
Age	1.000	(0.976, 1.025)	0.983
Race	0.895	(0.427, 1.814)	0.763

Table 9. Log-Rank Test Results for Objective #3.

Predictor	χ^2 p-value
BL_CKD	< 0.0001
Sex	0.999
Race	0.5
History of smoking	0.1
History of MI	0.8
Hypertension	0.3

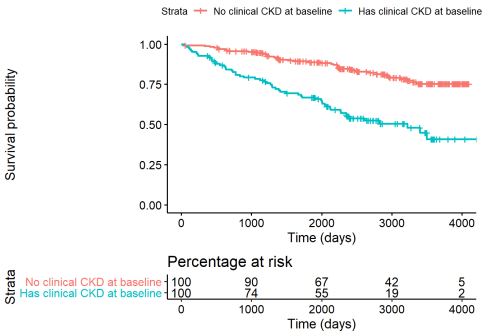
Table 10. Cox Proportional Hazards Model for Objective #4.

Predictor	$\hat{\beta}$	exp($\hat{\beta}$)	se($\hat{\beta}$)	z-score	p-value
BL_CKD	0.15	1.16	0.10	1.49	0.14
suPAR	0.00	1.00	0.00	1.06	0.29
EverMI	-0.11	0.89	0.19	-0.60	0.55
BMI	0.00	1.00	0.01	0.29	0.77
DM	-0.21	0.81	0.18	-1.17	0.24
HTN	-0.05	0.96	0.19	-0.24	0.81
EverSmoked	0.22	1.25	0.16	1.35	0.18
Dyslipidemia	-0.05	0.95	0.18	-0.27	0.79

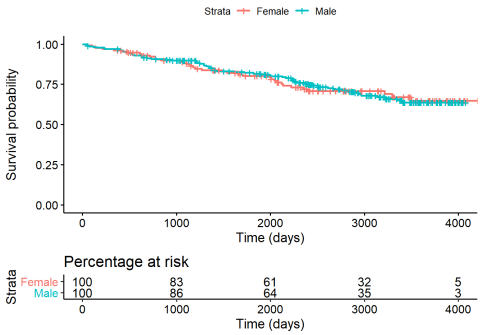
Table 11. Pearson's Correlation Test Results for Objective #4.

Predictor	Pearson's Correlation	p-value
BL_CKD	0.96	0.34
suPAR	-0.35	0.72
EverMI	1.01	0.32
BMI	-0.2	0.84
DM	-0.64	0.52
HTN	0.26	0.80
EverSmoked	0.3	0.76
Dyslipidemia	-1.66	0.10

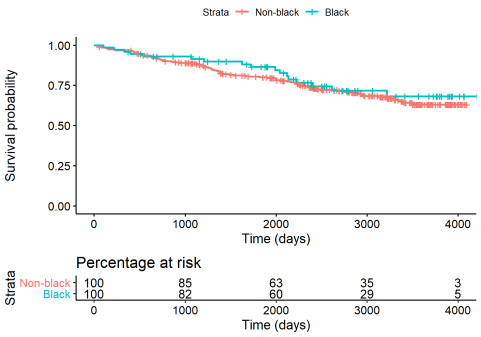
Figure 1. Kaplan-Meier Survival Curves for Time to Death Stratified by Six Covariates.



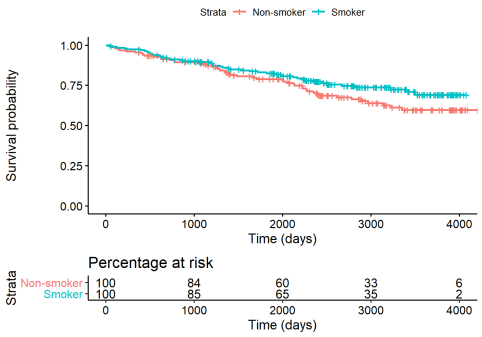
(a) Survival curves by clinical CKD at baseline.



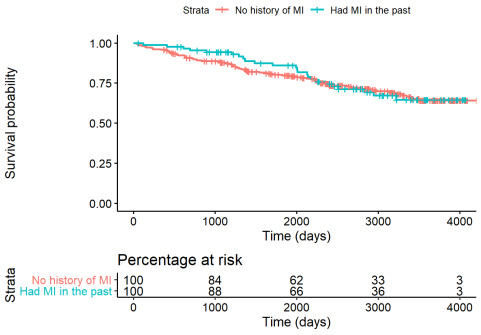
(b) Survival curves by sex.



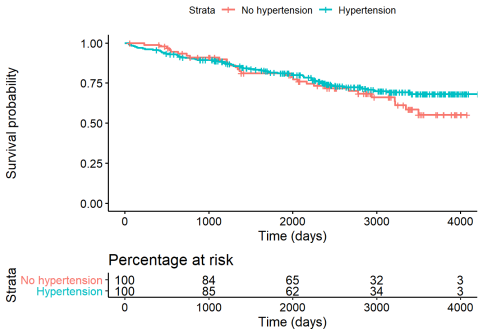
(c) Survival curves by race.



(d) Survival curves by smoking status.



(e) Survival curves by history of MI.



(f) Survival curves by hypertension.